PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2002-108382
(43)Date of publication of application : 10.04.2002

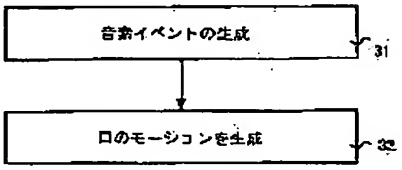
(51)Int.Cl. G10L 15/00
G06T 13/00
G06T 15/70
G10K 15/02
G10L 13/00
G10L 21/06
G10L 15/22

(21)Application number : 2000-294151
(22)Date of filing : 27.09.2000

(71)Applicant : SONY CORP
(72)Inventor : OTO YASUNORI
UEDA YUICHI

(54) ANIMATION METHOD AND DEVICE FOR PERFORMING LIP SINCHRONIZATION

(57)Abstract:
PROBLEM TO BE SOLVED: To provide animation technique capable of automatically generating an animation automatically including a mouth part according to voice data, etc.
SOLUTION: Voice data are received and a phoneme analysis is taken to generate phoneme events along a time base (31). A mouth-shape animation corresponding to one or more phonemes is weighted and added to generate an animation of a distinctive or vague mouth shape corresponding to an articulate or obscure phoneme (32). For a phoneme whose mouth movement is earlier than pronunciation, the phoneme event is made earlier than voice data timing.



LEGAL STATUS
[Date of request for examination]
[Date of sending the examiner's decision of rejection]
[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]
[Date of final disposal for application]
[Patent number]
[Date of registration]
[Number of appeal against examiner's decision of rejection]
[Date of requesting appeal against examiner's decision of rejection]
[Date of extinction of right]

(19)日本国特許庁（ＪＰ）　(12)公開特許公報（Ａ）　(11)特許出願公開番号
特開2002－108382
（P2002－108382A）
(43)公開日　平成14年４月10日(2002.4.10)

(51)Int.Cl.⁷	識別記号	ＦＩ		テーマコード（参考）
Ｇ１０Ｌ　15/00		Ｇ０６Ｔ　13/00	Ｂ	５Ｂ０５０
Ｇ０６Ｔ　13/00		15/70	Ｂ	５Ｄ０１５
15/70		Ｇ１０Ｋ　15/02		５Ｄ０４５
Ｇ１０Ｋ　15/02		Ｇ１０Ｌ　3/00	５５１Ｈ	
Ｇ１０Ｌ　13/00			Ｑ	

審査請求　未請求　請求項の数11　ＯＬ　（全９頁）　最終頁に続く

(21)出願番号　特願2000－294151(P2000－294151)

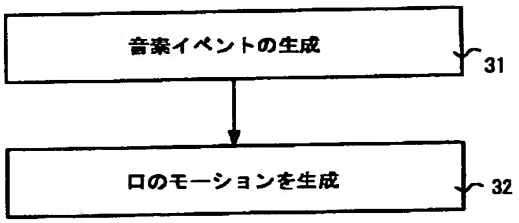
(22)出願日　平成12年９月27日(2000.9.27)

(71)出願人　000002185
ソニー株式会社
東京都品川区北品川６丁目７番35号
(72)発明者　大戸　康紀
東京都品川区東五反田１丁目14番10号　株式会社ソニー木原研究所内
(72)発明者　上田　裕一
東京都品川区東五反田１丁目14番10号　株式会社ソニー木原研究所内
(74)代理人　100101801
弁理士　山田　英治　（外２名）

最終頁に続く

(54)【発明の名称】　リップシンクを行うアニメーション方法および装置

(57)【要約】
【課題】　リップシンク・アニメーションを簡易に生成する。
【解決手段】　音声データを受けとって音素解析を行い、時間軸に沿って音素イベントを生成する（３１）。１または複数の音素に対応した口形状アニメーションを重み付け加算して明瞭な、あるいは、あいまいな音素に対応した明瞭な口形状、あるいは、あいまいな口形状のアニメーションを生成する（３２）。口の動きのほうが発声より早い音素の場合には音素イベントを音声データタイミングより早めにする。



【特許請求の範囲】
【請求項１】　音声データから音素データを生成するステップと、
上記音素データに基づいて二次元ないし三次元のモデルの口の部分のアニメーションを生成するステップとを有することを特徴とするアニメーション生成方法。
【請求項２】　上記口の部分のアニメーションを生成するステップは、音素データの組合せによりあいまいな音声を表し、この組合せに含まれる音素データにそれぞれ対応する口の部分のアニメーションを重み付け合成する請求項１記載のアニメーション生成方法。
【請求項３】　所定の音素データについては、音素データのタイミングより、対応する口の部分のアニメーションの生成のタイミングが所定時間早くする請求項１記載のアニメーション生成方法。
【請求項４】　音素データの生成を有限状態オートマトンを参照して行い音素間の遷移に所定の制約を設ける請求項１記載のアニメーション生成方法。
【請求項５】　音声データから音素データを生成するステップと、
上記音素データに基づいて二次元ないし三次元のモデルの口の部分のアニメーションを生成するステップと、
上記音素データを基準にして上記音声データと上記アニメーションとを同期させて再生するステップとを有することを特徴とするアニメーション生成方法。
【請求項６】　音素データに基づいて二次元ないし三次元のモデルの口の部分のアニメーションを生成し、さらに、音素データの組合せによりあいまいな音声を表し、この組合せに含まれる音素データにそれぞれ対応する口の部分のアニメーションを重み付け合成することを特徴とするアニメーション生成方法。
【請求項７】　音素データに基づいて二次元ないし三次元のモデルの口の部分のアニメーションを生成し、さらに、所定の音素データについては、音素データのタイミングより、対応する口の部分のアニメーションの生成のタイミングが所定時間早くすることを特徴とするアニメーション生成方法。
【請求項８】　音素データの組合せによりあいまいな音声を表し、この組合せに含まれる音素データにそれぞれ対応する口の部分のアニメーションを重み付け合成する請求項７記載のアニメーション生成方法。
【請求項９】　音声データから音素データを生成する手段と、
上記音素データに基づいて二次元ないし三次元のモデルの口の部分のアニメーションを生成する手段とを有することを特徴とするアニメーション生成装置。
【請求項１０】　音声データから音素データを生成する手段と、
上記音素データに基づいて二次元ないし三次元のモデルの口の部分のアニメーションを生成する手段と、
上記音素データを基準にして上記音声データと上記アニメーションとを同期させて再生する手段とを有することを特徴とするアニメーション生成装置。
【請求項１１】　音声データから音素データを生成するステップと、
上記音素データに基づいて二次元ないし三次元のモデルの口の部分のアニメーションを生成するステップと、
上記音素データを基準にして上記音声データと上記アニメーションとを同期させて再生するステップとをコンピュータに実行させるために用いるコンピュータプログラムをコンピュータ読取り可能に記録するアニメーション生成用の記録媒体。
【発明の詳細な説明】
【０００１】
【発明の属する技術分野】本発明は、二次元あるいは三次元形状のモデルにおいて、音声データに同期して、口となる部分のアニメーションを生成する技術に関する。本発明の技術はアニメーションの作成時あるいは実行時に使用され、ゲームや映像コンテンツなどで活用される。
【０００２】
【従来の技術】従来のリップシンクアニメーションでは、音声データを記録した後、そのデータに合わせてリップアニメーションを作成する必要があった。もしくは、予めあたりをつけておいた口のアニメーションに対して音声データを当てていく必要があった。
【０００３】
【発明が解決しようとする課題】従来のリップシンクアニメーションでは、音声データを見ながらアニメーションを作成していく場合、かなりの入力が必要であった。また、この場合、音声データとアニメーションとを分離することが難しく、リップアニメーションの再利用性もほとんどなかった。また、音声データを追加する場合、新たにリップアニメーションを作成したり、前に作成したアニメーションを修正する必要があった。
【０００４】また、アテレコによって音声データをアニメーションに付加していく場、アニメーションと音声の間に明確な対応関係がなく、声優の経験によるところが大きかった。
【０００５】本発明は上述の事情を考慮してなされたものであり、音声データ等に基づいて自動的に口の部分を含むアニメーションを自動的に生成することができるアニメーション技術を提供することを目的としている。
【０００６】
【課題を解決するための手段】本発明によれば、上述の目的を達成するために、特許請求の範囲に記載のとおりの構成を採用している。ここでは、本発明について若干補充的な説明を行う。
【０００７】本発明の原理的な構成例によれば、音声データの音素解析を行った後、解析で得られた音素の種類

に対する重みを計算し、あいまいな口形状を示すデータを作成する。次に、音素間の形状変化アニメーションの作成を行い、音声データに対してイベント作成した口形状アニメーションをマッピングする。この際、破裂音（／ｐ／や／ｂ／）などにおいて口形状の変化が音素の発音以前に行われるため、イベントを前方に修正する。口形状のアニメーションを行う時点で、音声データを再生し、これと同時に、イベントに基づいてアニメーション合成を行うことにより、同期の取れたアニメーションの表示を行う。

【０００８】このような構成によれば、二次元あるいは三次元モデルのアニメーションの作成とは別に、音声データのみを後から独立して追加することが可能であり、モデル設計時においてリップシンクアニメーションを個々に作成する必要がなくなる。さらに、音声データに対して、インタラクティブにリップアニメーションを生成することができる。また、／ａ／と／ｅ／の中間などのあいまいな発音に対して、あいまいな口の動きを実現することができる。また、子音に対して、有声音と無声音（／ｚ／と／ｓ／なと）や、破裂音と摩擦音（／ｐ／と／ｚ／）など、発音状態として同時にその状態を取り得ない場合や、音素間の連続性として、／ｙ，ａ／、／ｙ，ｙ／、／ｙ，ｏ／などが許されるのに対して、／ｙ，ｉ／、／ｙ，ｅ／が許されないなどといった制約を盛り込むことが可能になる。

【０００９】なお、音素としては、子音、母音単位であつかってもよいし、より大きな単位例えば音節のシンボルとして扱ってもよい。要するに口形状とマッピング可能な単位のものであればどのような音素を用いてもよい。

【００１０】また、本発明は方法としても装置としても実現可能である。また、そのような方法をコンピュータで実現するために用いるプログラムを記録したコンピュータ読取り可能な記録媒体も本発明の技術的な範囲に含まれる。

【００１１】

【発明の実施の形態】ここでは、本発明におけるリップシンクアニメーション生成およびその表示を実現するための装置の説明をした後、音声データの解析方法と、あいまいな発音に対する、口のあいまいなアニメーションの生成方法について説明を行う。

【００１２】図１は、本発明の実施例のアニメーション生成装置を全体として示しており、この図において、アニメーション生成装置１は、音素イベント生成部（アプリケーション）２、アニメーション処理部（アプリケーション）３、オペレーティングシステム４、入力装置５、出力装置６、その他のリソース（ハードウェア、ソフトウェア）７等を含んで構成されている。アニメーション生成装置１は、実際には、ゲーム機、マルチメディアキオスク、パーソナルコンピュータ等に実装される。アニメーション編集装置として構成してもよい。オペレーティングシステム４は、実装環境に応じたものであり、パーソナルコンピュータ用の汎用のオペレーティングシステムでもよいし、機器独自の組み込みオペレーティングシステムでもよい。音素イベント生成部２は、入力音声データを解析してアニメーション生成の音素イベントを生成するものである。詳細については図４を参照して後に説明する。アニメーション処理部３は、音素イベント生成部２から音素イベントを受けとってリップシンクのアニメーションを合成して画像データを生成するものである。図示しないが、画像生成の一部を、専用のハードウェアを用いて行ってもよい。

【００１３】図２は、図１の音素イベントおよびアニメーション処理の関係を説明するものである。図１において、音声データを解析して音素データ２１（「ｋｏ」、「ｎｎ」、「ｎｉ」、「ｔｉ」、「ｗａ」）が生成され、音素データの出現タイミングに応じて対応するアニメーション２２、２３、２４が生成される。各アニメーション２２、２３、２４は音素に応じた口形状をしており、音素イベントの出現に応じて対応するアニメーションを起動することによりリップシンクを実現できる。

【００１４】図３は、本実施例の動作を概要を示しており、この図に示すように、本実施例では、音声データに対する前処理を行って音素イベントを生成し（３１）、この後、音声データに同期した口のアニメーションを行っている（３２）。もちろん、音声データに対する前処理は逐次処理が可能であり、音声データをストリーミング入力しながら、アニメーションを生成することも可能となっている。これらの処理３１および３２は図１の音素イベント生成部２およびアニメーション処理部３にそれぞれ対応しており、それぞれ後に詳述する。

【００１５】図４は、音声データに対して行う前処理の流れを示す。図４において、音声データをシステムに入力し（４０１）、音素解析を行った後（４０２）、状態遷移を通すことによってその候補を制限する（４０３）。同時に複数の音素が候補として残る場合には、これによって得られた複数の音素分に対して口形状の重み割合を計算し、口形状の合成情報を作成する（４０４）。この後、音素遷移におけるアニメーションを作成し（４０５）、音素データに対してイベントとアニメーションの登録を行う（４０６）。

【００１６】次に、これらの処理について一つずつ説明を行う。

【００１７】図５に音声データの入力ソース（５０６）の一例を示す。インターネットにおけるストリーミングデータ（５０１）や放送（５０２）、マイクによる直接入力（５０３）や、ＣＤ（５０４）、ＭＤ（５０５）などの記録媒体からの入力が可能となっている。

【００１８】次に、音素解析について説明する。最初に図６に示すように音声の波形データ（６０２）におい

て、ゼロに交わる点（６０１）を抽出し、その周期性を調べる。子音が過渡的であるのに対し、母音部における周期性が揃っていることから、解析フレームを適宜作成していくことにより、母音・子音の分離と、スペクトル解析の精度を向上させることができる。ここで、６０３は波形の周期を示している。また６０４で示す部分は、波形の周期が過渡的であることから子音フレームとして判断されている。また、６０５で示す部分は波形の周期が揃っていることから、母音フレームと判断されている。

【００１９】次に、図7に示すように、入力された音声データ（ａ）において、図６で判断した解析フレームごとに周波数スペクトル（ｂ）、（ｃ）を求める。

【００２０】なお、周波数スペクトルの求め方としてはＦＥＴ（高速フーリエ変換）やＭＥＭ（最大エントロピー法）などがある。これらの手法は周知であるのでとくに説明は行わない。詳細については例えば「時系列解析プログラム」（北川源四郎著、岩波書店発行）を参照されたい。

【００２１】取得した周波数スペクトルはいくつかの主要となる周波数成分を持っており、図８（ａ）に示すように、低周波数側から、第一フォルマント（８０１）、第二フォルマント（８０２）、第三フォルマント（８０３）と言われている。本実施例ではこれらのフォルマント間の関係を用いて音素解析を行う。なお、フォルマントに関しては、「フーリエの冒険」（トランスナショナルカレッジオブブックス編集、ヒッポファミリークラブ発行）等を参照されたい。

【００２２】ここで、リップシンクアニメーションにおいては、同一モデルに対して一人の音声データ提供者が担当することが普通であり、この実施例では、特定話者を想定することにする。なお、不特定話者の場合に対しても、音素解析の手法が変化するだけであり、全体としては同一の構成となる。

【００２３】そして、特定話者を想定することから、図８（ｂ）に示すように、フォルマントと音素の関係についてキャリブレーションを行っておく。ここで、８０４～８０８は、各母音の位置を示している。また、これは話者の音質が変化しない限りにおいて、再度取り直す必要がない。

【００２４】図９（ａ）に示すように、音声データに対して設定した解析フレーム毎のフォルマントデータ９１を取得し、図８（ｂ）において行った、キャリブレーションしたフォルマント位置との関係を取得する。

【００２５】次に、図１０に示すような音素間の状態遷移を考慮し、音素候補の絞込みを行う。図１０（ａ）の状態遷移では、１００１は音素／ｋ／を表し、すべての母音（図１００２～１００６）への変化が可能であることを示している。一方、図１０（ｂ）の状態遷移では、１００７は音素／ｙ／を表し、／ａ／（１００８）、／ｕ／（１００９）、／ｏ／（１０１０）の母音へのみの状態遷移が許されていることを示している。

【００２６】また、図１１における、図１１０１から１１０２の遷移と、１１０３から１１０４への遷移によって生成されるアニメーション（１１０５から１１０６への変化）が同じものについては、一つにまとめて扱うことができる。

【００２７】次に、図１２に示すように、音素候補（１２０２，１２０３，１２０５，１２０６，１２０７）におけるフォルマント位置と計測されたフォルマント位置（図１２０４）の距離を計算し、各音素候補毎の重みを計算する。重み計算例を図１２０１に示す。この際、人間の耳は実際の周波数より、オクターブとして認識することから、各成分に対して対数を取った上で扱っている。

【００２８】また、図１３に示すように、第三フォルマントを用いる場合も、距離の計算方法（１３０１）が変化する以外は、音素候補（１３０２，１３０３，１３０５，１３０６）におけるフォルマント位置と計測されたフォルマント位置（１３０４）から、同様に計算することができる。

【００２９】次に、図１４に示すように、各音素フレームにおける口形状を各音素に対応する口形状の重み付け合成として表現する。ここで、１４０１～１４０３は各フレームにおける音素毎の重み配分を示している。また、図１４０４～１４０６は、各々の重み付け合成によって生成される口形状を示している。

【００３０】また、図１５に示すように、子音発音時において、前後の口形状（１４０１、１５０３）から、音素間のアニメーション（１５０２）を作成する。このとき、破裂音など、音素が実際に発音される前に口形状が変化しているものに関しては、イベントを前に移動（１５０４）しておく。また、アニメーション時間として、移動後のフレーム時間を用いるとする。ここで、図１５０５は／ｐ／の発音イベントを示しており、この時、アニメーション（１５０２）に割り振っておく。また、図１５０６は／ａ／の発音イベントを示している。

【００３１】図１６に示すように、図１５において作成した音素アニメーションとイベントデータの発生（１６０１）に従って、口形状のアニメーションを起動していく（１６０２）。次に、移動中のアニメーションの合成（１６０３）したあと、画面への表示（１６０４）を行う。

【００３２】なお、アニメーションの合成に関しては、例えば、複数の基本のアニメーションを所定の係数で線形加算して目的のアニメーションを合成することができる。線形加算による合成する手法は、周知の手法を採用できる。例えば、特開平２０００—１１１９９号公報「アニメーションの自動生成方法」を採用してもよい。

【００３３】アニメーション表示の様子を図１７に示

す。まず、音声データ（１７１３）の再生を行い、これにマッピングされているイベント（１７１０，１７１１，１７１２）とアニメーションを起動していく。次に、起動したアニメーションの合成を行い（１７０８）最終的に口形状のアニメーションを表示する（１７０１，１７０２，１７０３）。ここで、１７０９はイベントデータを示しており、また１７０５，１７０６，１７０７は各アニメーションを示している。

【００３４】なお、ストリーミングによって音声データが配信される場合には、音素解析が行われている時間分バッファリング（遅れ）が生じるが、これは計算機の能力向上に伴って遅れ時間を短くすることが可能である。

【００３５】

【発明の効果】前記のように、本発明によれば、音素データに対するリップシンクアニメーションを予め作成しておくのではなく、音声データから適宜、合成によって生成するために、後から音声データを作成し、追加することが容易になる。また、音素解析におけるあいまいさに対してあいまいな口表現を行うという意味を付けることによって、より自然な口の表現が可能になる。

【図面の簡単な説明】

【図１】　本発明の実施例の実現例を全体として示す図である。

【図２】　上述実施例の概要を説明する図である。

【図３】　上実施例の動作の流れの概要を示す図である。

【図４】　上述実施例の音素イベントの生成を説明する図である。

【図５】　上述実施例における音声データの入力ソースを説明する図である。

【図６】　上述実施例における音素解析用の解析フレームを説明する図である。

【図７】　上述実施例の解析フレーム単位の周波数スペクトルの取得を説明する図である。

【図８】　周波数スペクトルのフォルマント位置のキャリブレーションを説明する図である。

【図９】　解析フレーム単位でフォルマントを取り出すことを説明する図である。

【図１０】　音素間の状態遷移図を説明する図である。

【図１１】　同じ口のアニメーションとして集約できる状態遷移を一つにまとめる態様を説明する図である。

【図１２】　計測されたフォルマントと、キャリブレーションとの関係から重みを計算する態様を説明する図である。

【図１３】　子音に関して重み計算を行う態様を説明する図である。

【図１４】　音素に対応した口の形状の合成状態としてあいまいな口を表現することを示す図である。

【図１５】　音素間における口形状のアニメーションを作成することを示す図である。

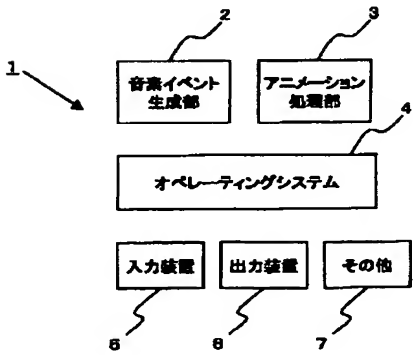
【図１６】　くりアニメーションの合成と表示の流れを説明する図である。

【図１７】　音声データの再生に合わせて口形状を表すアニメーションを合成していく様子を説明する図である。

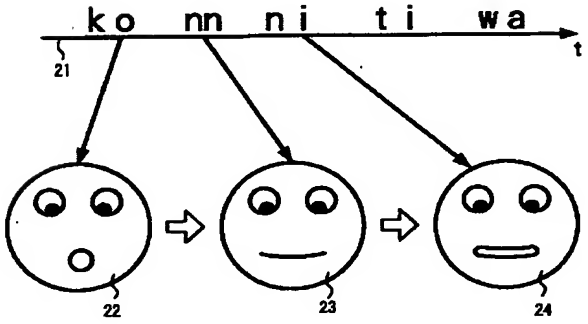
【符号の説明】

1	アニメーション生成装置
2	音素イベント生成部
3	アニメーション処理部
4	オペレーティングシステム
5	入力装置
6	出力装置

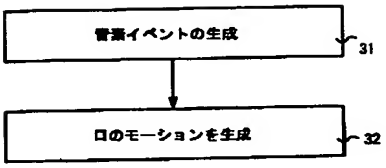
【図１】



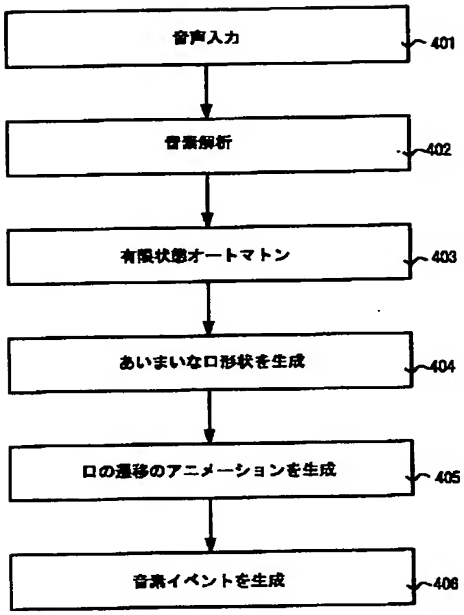
【図２】



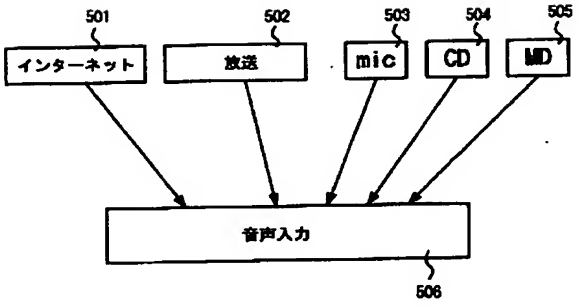
【図３】


音素イベントの生成
31
口のモーションを生成
32

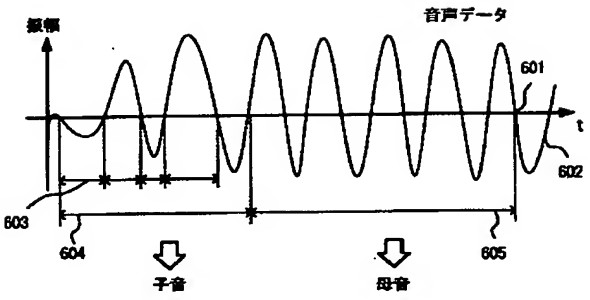
【図４】


音声入力
401
音素解析
402
有限状態オートマトン
403
あいまいな口形状を生成
404
口の遷移のアニメーションを生成
405
音素イベントを生成
406

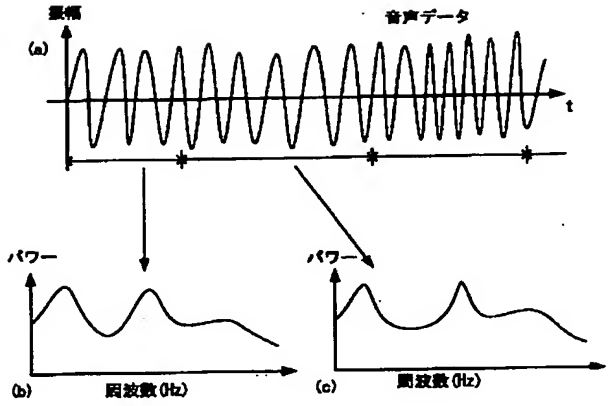
【図５】


501
502
503
504
505
インターネット
放送
mic
CD
MD
音声入力
506

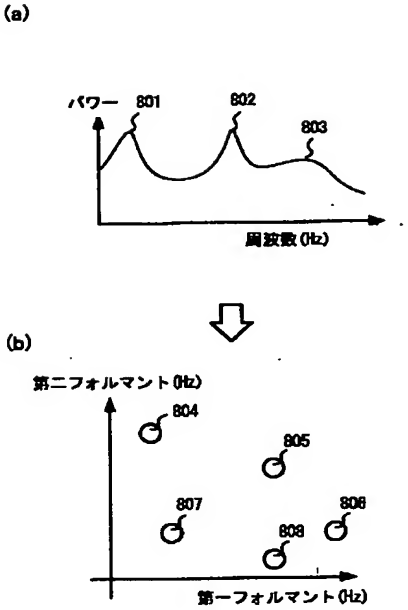
【図６】


振幅
音声データ
601
t
602
603
604
605
子音
母音

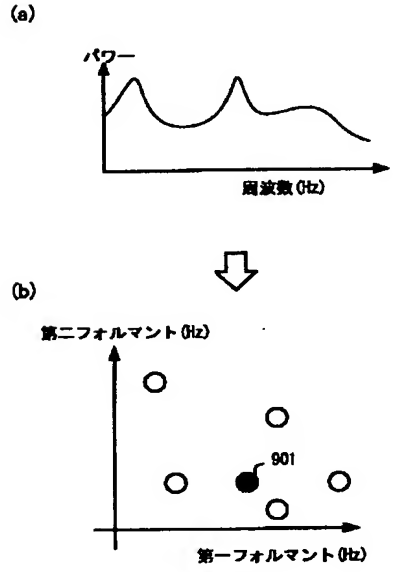
【図７】



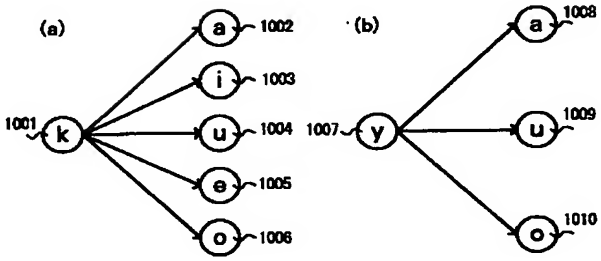
【図８】



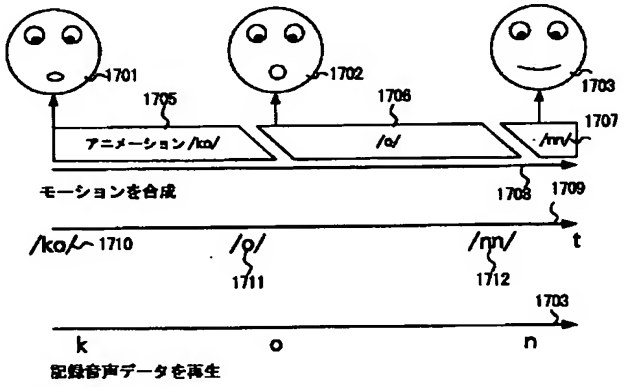
【図９】



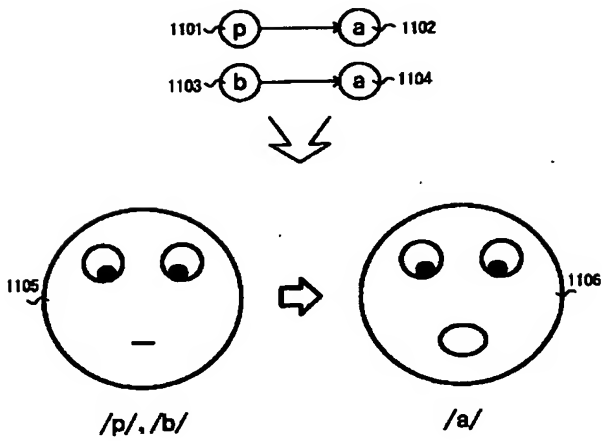
【図１０】



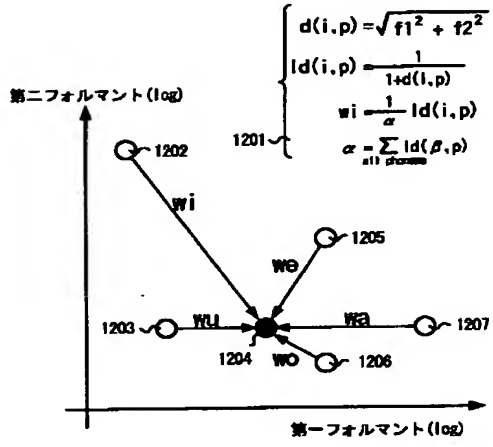
【図１７】



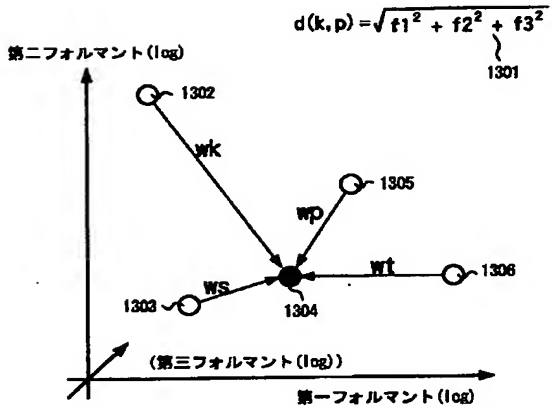
【図１１】



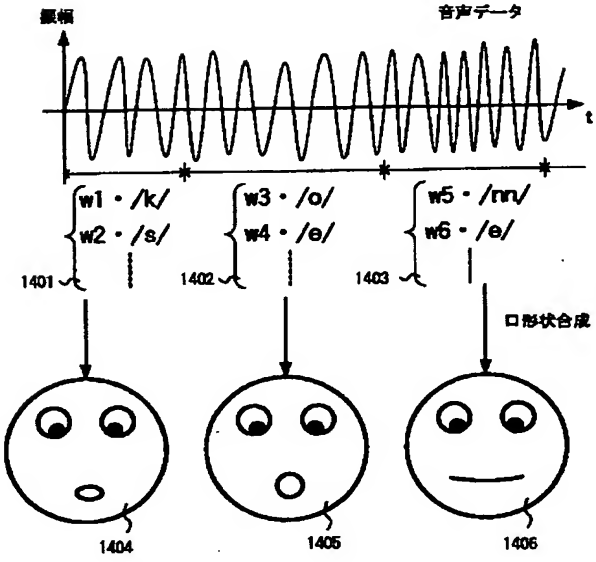
【図１２】



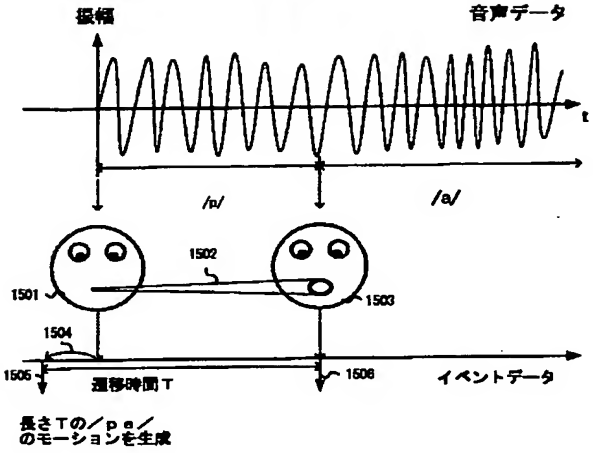
【図１３】



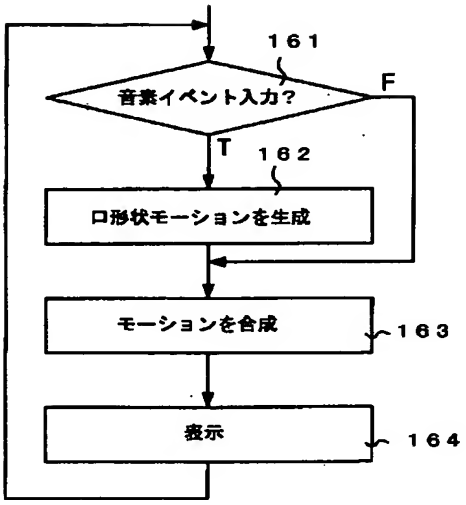
【図１４】



【図１５】



【図１６】



フロントページの続き

(51) Int. Cl.'	識別記号	ＦＩ		テーマコード' (参考)
Ｇ１０Ｌ　21/06		Ｇ１０Ｌ　3/00	Ｓ	
15/22			５６１Ｃ	

Ｆターム(参考)　5B050 BA08 BA12 EA19 EA24 FA10
5D015 AA05 BB02 CC03 CC04 DD02
KK01 LL12
5D045 AB01 AB11